Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–27

DOI: 10.1177/0962280215608213
smm.sagepub.com

SAGE

Fast and highly efficient pseudo-likelihood methodology for large and complex ordinal data

Anna Ivanova,[1] Geert Molenberghs[1,2] and Geert Verbeke[1,2]

Abstract

In longitudinal studies, continuous, binary, categorical, and survival outcomes are often jointly collected, possibly with some observations missing. However, when it comes to modeling responses, the ordinal ones have received less attention in the literature. In a longitudinal or hierarchical context, the univariate proportional odds mixed model (POMM) can be regarded as an instance of the generalized linear mixed model (GLMM). When the response of the joint multivariate model encompass ordinal responses, the complexity further increases. An additional problem of model fitting is the size of the collected data. Pseudo-likelihood based methods for pairwise fitting, for partitioned samples and, as introduced in this paper, pairwise fitting within partitioned samples allow joint modeling of even larger numbers of responses. We show that that pseudo-likelihood methodology allows for highly efficient and fast inferences in high-dimensional large datasets.

Keywords

Generalized linear mixed model, proportional odds mixed model, joint modeling, pseudo-likelihood, pairwise fitting, sample partition, asymptotic relative efficiency, reduced computation time

1 Introduction

Many statistical models have been developed for analyzing longitudinal data. Most of them are limited to the analysis of a single outcome, measured repeatedly over time. The random-effect approach has been very popular for several decades. The introduction of linear mixed models (LMM) for continuous data by Laird and Ware[1] was extended to generalized linear mixed models (GLMM) for non-continuous data by Breslow and Clayton,[2] Wolfinger and O'Connell,[3] and Engel and Keen.[4] Multinomial (nominal and ordinal) data can be regarded as a special case of

[1]I-BioStat, KU Leuven, University of Leuven, Leuven, Belgium
[2]I-BioStat, Universiteit Hasselt, Hasselt, Belgium

Corresponding author:
Anna Ivanova, I-BioStat, KU Leuven, University of Leuven, Leuven, Belgium.
Email: anna.ivanova@lstat.kuleuven.be

non-continuous data. Some examples of models for the latter data type can be found in the literatures,[5–7] such as the proportional odds mixed model (POMM) described by Agresti and Lang.[5]

Very often the collected outcomes are considered not as separate endpoints, but as components in a joint one. For example, in a diabetes study, one can model the body mass index and cholesterol level collected repeatedly from the same patient. Joint modeling will be the preferable technique because every outcome may have its own covariate structure as random effect, and at the same time the association between the outcomes can be captured in terms of the correlation between random effects.

Another feature of longitudional studies is the diversity in data type: continuous, binary, ordinal, survival data could be jointly collected and chosen to be jointly modelled. In the literature, a wide variety of joint modeling techniques can be found. For example, Morrell et al.[8] considered three continuous responses for screening prostate data, and further used the outcome for classification purposes. Gueorguieva[9] considered the joint modeling of continuous-binary measures in a toxicity study of pregnant/non-pregnant mice. Iddi and Molenberghs[10] considered the joint modeling of a continuous visual activity outcome and a binary vision-loss outcome in an age-related macular degeneration study. These authors also considered two binary longitudinal outcomes: the number of positive HCV and HIV cases in serological data. There also exist many approaches for the joint modeling of longitudinal and survival data (e.g. Rizopoulos[11]). For an excellent relatively early review, we refer to the work of Tsiatis and Davidian.[12] Molenberghs and Verbeke[13] discuss a number of techniques that jointly model outcomes of different types. But the advent of larger data storage facilities led to an additional complication: high-dimensional data, calling for a solution for large and/or complex data.

To address the problem of computational complexity when jointly modeling random effects in the high-dimensional case, Fieuws and Verbeke[14] suggested a bivariate pairwise approach using pseudo-likelihood. The method was illustrated in a joint analysis of 22 longitudinally measured outcomes. Since then, the technique has been applied to binary data[15] and to a combination of continuous and binary data.[16] This method is also reviewed by Molenberghs and Verbeke[13] (chap. 24). Molenberghs et al.[17] went on to propose a method to solve the problem of large data by partitioning into subsamples that are analyzed separately and by combining the obtained inferences into a single one. Two different scenarios were considered: independent and dependent partitioning. It was shown that, to achieve high relative efficiency for small samples, the data should be divided into subsamples such that the size of subsamples is much larger than the number of subsamples. In general, the number of subsamples for splitting depends on the pertinent setting of the modeling such as length of the sequence of longitudinal data and the complexity of the model such as the number of fixed effects on the one hand and the dimensionality of the random effects on the other hand.

In Vasdekis et al.,[18] a weighted pairwise likelihood estimation method was proposed based on estimates obtained from separate maximizations of marginal pairwise likelihoods. The weighted estimator is found to be more efficient than the one that assumes all weights to be equal.

In this work, we investigate the performance of both pairwise- and partition fitting on ordinal data. In addition, we develop a new approach in multivariate joint modeling based on pseudo-likelihood: we introduce the pairwise fitting within independent subsamples and combine the obtained inferences. The proposed method is based on the asymptotic inference of the parameter estimates and can be applied to every type of data. When modeling hierarchical ordinal responses, the complexity increases dramatically. Hence, the value of our method becomes even more significant.

We will compare all three methods with the full likelihood approach. The main emphasis of this paper is to illustrate that appropriate partitioning of data and models, in combination with

pseudo-likelihood methodology, makes fitting complex models to large data sets feasible, which otherwise would not be feasible.

The paper is organized as follows. The motivating data are introduced in Section 2 and their analysis discussed in Section 6. The theoretical concepts behind the proportional odds mixed model (POMM) and short overview of the classical approach for joint modeling are given in Sections 3 and 4, respectively. In Section 5, the existing pseudo-likelihood methodology will be reviewed and the new combined method introduced. Concluding remarks are given in Section 7.

2 Motivating case study

In Belgium, the diabetes project was conducted from January 2005 until December 2006, with the aim to study the effect of implementing a structured model for chronic diabetes care based on the patients' clinical outcomes. General practitioners (GP's) were offered assistance and could redirect patients to the diabetes care team, consisting of a nurse educator, a dietician, an ophthalmologist, and an internal medicine doctor. A total of 120 GP's and 2495 patients took part in the study.

During the project, several outcomes useful to evaluate how well diabetes is controlled were measured, at the moment the program was initiated (time T_0) and one year later (T_1). The most important outcomes were LDL-cholesterol (low-density lipoprotein cholesterol, mg/dl), HbA1c (glycosylated hemoglobin, %) and SBP (systolic blood pressure, mmHg). Furthermore, experts specified cut-off values to divide every outcome into more than two target groups (*multiple targets*). These clinical targets are of major scientific interest. For example, the values of HbA1C were divided into three groups: (1) $<7\%$, (2) $>7\%$ and $<8\%$, and (3) $\geq 8\%$. Hence, when the patient moves to a lower HBA1C group, as a result of the treatment, then this can be regarded as an improvement of his/her physical condition. The target groups for LDL-cholesterol and SBP were defined in a similar way. If one is interested in verifying whether a new care program simultaneously improves the targets of HbA1C, LDL-cholestrol and SBP, then the *multiple targets* of each response can no longer be regarded as separate outcomes, but as part of a joint trial endpoint. The data are discussed in detail in Borgermans et al.[19]

In this paper, we will study modeling of the joint trial endpoint, in order to be able to study the association between separate components, and how it evolves over time, as well as to explore some demographical and disease-related characterictics of the patient. More details about the modeling will be discussed in the methodology section. Because of missing values in one or more of the covariates in the model, the data were reduced to 2259 patients.

Time-point-specific descriptions of multiple targets for LDL-cholesterol, HbA1C, and SBP are listed in Table 1. Table 2 shows the number of patients having 0, 1, 2, or 3 components of the joint trial endpoint for the two time-points. From this we observe that 1659 patients have measurements for all three components for both time-points. Among the incompleters, three patients had only SBP measurements and no measurements of LDL-Cholesterol and Hb1Ac.

3 Generalized linear mixed models

The principle of linear mixed models, introduced by Laird and Ware,[1] was reformulated to non-continuous data (generelized linear mixed models, GLMM) by Breslow and Clayton,[2] Wolfinger and O'Connell,[3] and Engel and Keen.[4] A nice overview of the existing GLMM with a lot of applications can be found in the book of Molenberghs and Verbeke.[13]

Assume that a longitudinal non-normal outcome can be appropriately modelled using a mixed model. For this outcome, let Y_{ij} denote the jth measurement for subject $i = 1, \ldots, N$, $j = 1, \ldots, n_i$.

Table 1. Diabetes data: number of observations with corresponding targets of the outcomes at two time-points.

		# Observations	
LDL-chol. Targets		T_0	T_1
1:	<100 mg/dl	819	1106
2:	≥ 100 mg/dl and < 115 mg/dl	381	312
3:	≥ 115 mg/dl and < 130 mg/dl	287	220
4:	≥ 130 mg/dl	485	250
Missing		287	371
HbA1C Targets		T_0	T_1
1:	$<7\%$	1201	1357
2:	$\geq 7\%$ and $<8\%$	604	474
3:	$\geq 8\%$	413	176
Missing		41	252
SBP targets		T_0	T_1
1:	≤ 130 mmHg	1103	1152
2:	> 130 mmHg and ≤ 140 mmHg	551	469
3:	> 140 mmHg and ≤ 160 mmHg	466	324
4:	> 160 mmHg	136	75
Missing		3	239

Table 2. Diabetes data: number of patients with number of available components for two time-points.

		T_1				
		0	1	2	3	Total
T_0	0	0	0	1	2	3
	1	2	2	5	12	21
	2	45	2	39	195	280
	3	179	16	101	1659	1955
	Total	226	19	146	1868	2259

The n_i measurements are grouped into a vector $\boldsymbol{Y}_i$. The GLMM assumes that conditionally on the random effects $\boldsymbol{b}_i$, with zero mean and covariance matrix D, the elements Y_{ij} of $\boldsymbol{Y}_i$ are independent, with densities belonging to the exponential family, i.e. of the form

$$f_i(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\xi}, \phi) = \exp\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\},$$

with

$$\eta[\psi'(\lambda_{ij})] = \eta(\mu_{ij}) = \eta[E(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{\xi})] = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i,$$

in which $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are k-dimensional and q-dimensional vectors with known covariate values, $\boldsymbol{\xi}$ a k-dimensional vector of unknown fixed regression coefficients, and ϕ a scale parameter.

In general, inference is based on the marginal model for $\boldsymbol{Y}_i$, which is obtained by integrating out the random effects

$$f_i(\boldsymbol{y}_i) = \int f_i(\boldsymbol{y}_i|\boldsymbol{b}_i) f(\boldsymbol{b}_i) \mathrm{d}\boldsymbol{b}_i. \tag{1}$$

In contrast to linear mixed models, the integral in (1) cannot be calculated analytically for most generalized linear mixed models, while existing analytical expressions tend to be cumbersome.[20,21] Hence, numerical approximations are needed (see Molenberghs and Verbeke[13]).

A special case of GLMM, of particular interest to this work, is the proportional odds mixed model (POMM) for ordinal outcomes. Let Y_{ij} be ordinal with values $r = 1, \ldots, R$. We first define R indicator variables as

$$W_{r,ij} = \begin{cases} 1 & \text{if } Y_{ij} = r, \\ 0 & \text{otherwise.} \end{cases}$$

Evidently, the R dummies are jointly redundant but any $R-1$ subset is not. Group the dummies into vectors $\boldsymbol{W}_{ij}$ for a specific subject i and occasion j, and further into $\boldsymbol{W}_i$ for all dummies across all occasions for subject i. We assume a multinomial distribution $\boldsymbol{W}_{ij} \sim$ multinomial (π_{ij}), with $\pi_{ij} = (\pi_{1,ij}, \ldots, \pi_{r,ij}, \ldots, \pi_{R,ij})$. The multinomial distribution at a given occasion is determined by the modeling choice made for the ordinal outcome. The probabilities can be written as

$$\pi_{r,ij} = \begin{cases} \kappa_{1,ij} & \text{if } r = 1, \\ \kappa_{r,ij} - \kappa_{r-1,ij} & \text{if } 1 < r < R, \\ 1 - \kappa_{R-1,ij} & \text{if } r = R \end{cases}$$

where, assuming proportional odds

$$\kappa_{r,ij} = \frac{\exp\left(\xi_{0r} + \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i\right)}{1 + \exp\left(\xi_{0r} + \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i\right)}.$$

Here, $\xi_{01} \leq \ldots \leq \xi_{0(R-1)}$ are intercepts, $\boldsymbol{\xi}$ fixed regression coefficients, $\boldsymbol{b}_i$ a vector of normally distributed random effects, and $\boldsymbol{x}_{ij}$ ($\boldsymbol{z}_{ij}$) the design vector for the fixed (random) effects at occasion j.

4 Joint models for multiple outcomes: classical approach

Consider L longitudinal outcomes, all of ordinal type. For each outcome, a POMM, as described in Section 3, can be specified. All L outcomes can be modelled jointly by defining a joint distribution of random effects. Now, $\boldsymbol{b}_i$ represents the vector with all random effects of all POMMs. Further, we will call this model *multivariate joint* POMM.

Assuming subjects to be independent, then it immediately follows from the independence of $\boldsymbol{Y}_{1,i}, \boldsymbol{Y}_{2,i}, \ldots, \boldsymbol{Y}_{L,i}$ conditional on $\boldsymbol{b}_i$, that the log-likelihood contribution for subject i to the full joint mixed model is given by

$$l_i(\boldsymbol{\theta}^*|\boldsymbol{y}_{1i}, \boldsymbol{y}_{2i}, \ldots, \boldsymbol{y}_{Li}) = \log\left(\int \prod_{l=1}^{L} f_{li}(\boldsymbol{y}_{li}|\boldsymbol{b}_i) f(\boldsymbol{b}_i) \mathrm{d}\boldsymbol{b}_i\right), \tag{2}$$

with $\boldsymbol{\theta}^*$ the vector of all parameters (fixed effects as well as covariance parameters). Except for special cases (e.g. with linear models), the integral in (2) cannot be calculated analytically and numerical approaches are needed. In this paper, we will use numerical integration, more specifically adaptive Gaussian quadrature, which has been implemented in the SAS procedure NLMIXED.[13,22,23] The potentially high dimension of the random effects in $\boldsymbol{b}_i$ rapidly leads to an increase in computation time because of the joint maximization of a large number of terms that need to be numerically evaluated.

5 Pseudo-likelihood methodology

5.1 Introduction

As mentioned in Section 4, in general, fitting joint GLMMs by maximizing likelihood becomes cumbersome and for large numbers of outcomes even infeasible. Instead, an alternative approach based on pseudo-likelihood methodology can be used, the principal idea of which is to replace the joint likelihood function by a function that is easier to maximize. This methodology makes use of estimating functions,[24] like many other methods (generalized estimating equations, partial likelihood, etc.). An important difference with, for example, partial likelihood is that pseudo-likelihood retains a fully parametric specification. Taking into account suitable regularity conditions, it can be shown that pseudo-likelihood maximization yields a consistent, asymptotically normally distributed estimator for the parameter vector.[25,26] In the book of Molenberghs and Verbeke[13] (chaps 9, 12, 21, 22, 24 and 25), the formal definition of pseudo-likelihood is given, its asymptotic properties and pseudo-likelihood inference considered, and a broad range of applications discussed in the case of marginal, conditional and subject-specific models, as well as joint modeling.

In the present paper, we will apply three different fitting techniques based on pseudo-likelihood: pairwise modeling, independent partitions modeling and, finally, the combined case: pairwise modeling of independent partitions. For each modeling case, a pseudo-likelihood function will be formulated and the method for estimating the parameters and the covariance matrix discussed.

5.2 Pairwise modeling

Applying the methodology of Fieuws and Verbeke,[14] we fit a bivariate joint POMM to each possible pair of longitudinal outcomes. Hence, for L longitudinal sequences $L(L-1)/2$ bivariate models will be fitted. Of course, in the pairwise approach, most of the parameters (but not all of them) will be estimated $L-1$ times: for example, consider the fixed effects in the model for the first longitudinal outcome, because the first outcome will be paired with the second, then with the third, and so on, up to the Lth outcome. In order to obtain a single estimate for the parameters in $\boldsymbol{\theta}^*$, an average overall different bivariate POMMs is taken.

The first formal step of this procedure is to maximize the log-likelihood of each bivariate model separately

$$l_{rs}(\boldsymbol{\theta}_{r,s}|\boldsymbol{y}_r,\boldsymbol{y}_s)=\sum_{i=1}^{N} l_{rsi}(\boldsymbol{\theta}_{r,s}|\boldsymbol{y}_{ri},\boldsymbol{y}_{si}),$$

where $r=1,\ldots,L-1$, and $s=r+1,\ldots L$, and N the number of subjects. As result, we obtain an estimate for $\boldsymbol{\theta}_{r,s}$, the vector of all parameters in a specific pair (r, s).

Let $\boldsymbol{\theta}$ be the vector contaning the parameters of all $L(L-1)/2$ bivariate models. The pseudo-likelihood then takes the following form

$$pl(\boldsymbol{\theta}) = l_{12}(\boldsymbol{\theta}_{1,2}|\boldsymbol{y}_1, \boldsymbol{y}_2) + l_{13}(\boldsymbol{\theta}_{1,3}|\boldsymbol{y}_1, \boldsymbol{y}_3) + \cdots + l_{(L-1)L}(\boldsymbol{\theta}_{L-1,L}|\boldsymbol{y}_{L-1}, \boldsymbol{y}_L).$$

Then, within the pseudo-likelihood framework, as was shown for the vector-valued parameter case by Arnold and Strauss[25] and by Geys et al.,[26] an asymptotic multivariate distribution for $\hat{\boldsymbol{\theta}}$ can be derived as follows

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, \boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1})$$

where $\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1}$ is a 'sandwich-type' robust variance estimator (see Appendix 1). Let $\hat{\boldsymbol{\theta}}^*$ be our vector of interest, the vector of the average over all available estimates. Then, to pass from the distribution of $\hat{\boldsymbol{\theta}}$ to $\hat{\boldsymbol{\theta}}^*$, Fieuws and Verbeke[14] used an appropriate linear combination matrix $\boldsymbol{A}$. Then, $\hat{\boldsymbol{\theta}}^* = \boldsymbol{A}'\hat{\boldsymbol{\theta}}$ and the pseudo-likelihood inferences for the elements of $\hat{\boldsymbol{\theta}}^*$ will be based on the following assymptotic distribution

$$\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) = \sqrt{N}(\boldsymbol{A}'\hat{\boldsymbol{\theta}} - \boldsymbol{A}'\boldsymbol{\theta}) \sim N(\mathbf{0}, \boldsymbol{A}'\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1}\boldsymbol{A}). \tag{3}$$

5.3 Partitioned samples

The idea of Fieuws and Verbeke,[14] reviewed in Section 5.2, was adopted by Molenberghs et al.[17] and applied to the case where partitioning of large data is appropriate for model fitting.

Consider a large sample, broken into $m = 1, \ldots, M$ independent subsamples, each of size n. Then, $N = M \cdot n$. (The extension to different subsample sizes is straightforward, then $N = \sum_{m=1}^{M} n_m$). In this case, the fist step is to maximize the log-likelihood for each subsample separately

$$l_m(\boldsymbol{\theta}_m|\boldsymbol{y}_{m,1}, \ldots, \boldsymbol{y}_{m,L}) = \sum_{i=1}^{n} l_{mi}(\boldsymbol{\theta}_m|\boldsymbol{y}_{m,1}, \ldots, \boldsymbol{y}_{m,L}).$$

The estimation of the parameters will be done by maximizing the following pseudo-likelihood function

$$pl(\boldsymbol{\theta}) = l_1(\boldsymbol{\theta}_1|\boldsymbol{y}_{1,1}, \ldots, \boldsymbol{y}_{1,L}) + l_2(\boldsymbol{\theta}_2|\boldsymbol{y}_{2,1}, \ldots, \boldsymbol{y}_{2,L}) + \cdots + l_M(\boldsymbol{\theta}_M|\boldsymbol{y}_{M,1}, \ldots, \boldsymbol{y}_{M,L})$$

Note that all $\boldsymbol{\theta}_m$ are equal to $\boldsymbol{\theta}^*$ and the parameter vector $\boldsymbol{\theta}$ from Section 5.2 takes the form $(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*, \ldots, \boldsymbol{\theta}^*)$. Then, the average estimator over all subsamples can be defined as follows

$$\hat{\boldsymbol{\theta}}^* = \frac{1}{M}\sum_{m=1}^{M} \hat{\boldsymbol{\theta}}_m. \tag{4}$$

Due to the fact that modeling was performed on independent subsamples, the mutual information between the subsamples will be zero. Hence, blocks $\boldsymbol{J}_m$ and $\boldsymbol{K}_m$ are identical up to the sign: $\boldsymbol{J}_m = -\boldsymbol{K}_m$ (see Appendix 2). Using these results, the approximated distribution in (3) can be simplified to

$$\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) = \sqrt{N}(\boldsymbol{A}'\hat{\boldsymbol{\theta}} - \boldsymbol{A}'\boldsymbol{\theta}) \sim N(\mathbf{0}, -M \cdot \boldsymbol{A}'\boldsymbol{J}^{-1}\boldsymbol{A}) \tag{5}$$

where

$$\boldsymbol{A} = \frac{1}{M}(\boldsymbol{I}, \boldsymbol{I}, \ldots, \boldsymbol{I}) \tag{6}$$

with $\boldsymbol{I}$ the identity matrix with linear dimensions equal to the length of vector $\boldsymbol{\theta}^*$. An alternative way to estimate the covariance matrix of $\hat{\boldsymbol{\theta}}^*$ is by using the observed information matrices that contain the second derivatives of the pseudo-likelihood

$$\sum(\hat{\boldsymbol{\theta}}^*) = \frac{1}{N}\left(-\frac{1}{M}\sum_{m=1}^{M}\widehat{\boldsymbol{J}_m}^{-1}\right). \tag{7}$$

It is important to note that this method is fully efficient in the sense that the Cramér-Rao lower bound is reached. This is due to the fact that, for every partition, a genuine likelihood was used for model fitting. This is true even though the obtained vector of estimates may differ from the one obtained when analyzing the outcomes in full.

Molenberghs et al.[17] also investigated the case of model fitting given dependent subsamples. Then, the asymptotic distribution of the vector $\hat{\boldsymbol{\theta}}^*$ will be derived using the general case, as in equation (3), with $\boldsymbol{A}$ as in equation (6). The reason behind it is that the off-diagonal blocks of $\boldsymbol{K}$, the information matrix with the products of the first derivatives, are generally non-zero.

5.4 Pairwise modeling of independent subsamples

In this paper, we will introduce a combined method: pairwise modeling of independent subsamples. Therefore, formally, we have to take the following steps: (1) divide the data into M independent subsamples, each of size n (or, in general, n_m), (2) apply pairwise fitting on each subsample separately, and (3) combine the results from the two previous steps in a single inference for vector $\hat{\boldsymbol{\theta}}^*$. Clearly, model fitting is in this case is based on the pseudo-likelihood function of the form

$$pl(\boldsymbol{\theta}) = \sum_{m=1}^{M}\sum_{r,s} l_{m,rs},$$

where

$$l_{m,rs}(\boldsymbol{\theta}_{m,rs}|\boldsymbol{y}_{m,r}, \boldsymbol{y}_{m,s}) = \sum_{i=1}^{n} l_{m,rsi}(\boldsymbol{\theta}_{m,rs}|\boldsymbol{y}_{m,ri}, \boldsymbol{y}_{m,si}),$$

and $r = 1, \ldots, L-1$, and $s = r+1, \ldots L$, $\boldsymbol{y}_{m,ri}$ and $\boldsymbol{y}_{m,si}$ are subvectors of outcomes r and s for subject i, the subject that is included in subsample m. Note that the independent subsamples contain different subjects.

For each of m independent subsamples (because of pairwise modeling), most parameters will be estimated $L-1$ times. Hence, in this case, matrix $\boldsymbol{A}$ will contain the weights over all partitions and all pairs. The asymptotic distribution of the vector of estimators $\hat{\boldsymbol{\theta}}^* = \boldsymbol{A}'\hat{\boldsymbol{\theta}}$, averaged over all pairs and partitions, will be a generalization of (3) and (5)

$$\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) = \sqrt{N}(\boldsymbol{A}'\hat{\boldsymbol{\theta}} - \boldsymbol{A}'\boldsymbol{\theta}) \sim N(\boldsymbol{0}, M \cdot \boldsymbol{A}'\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1}\boldsymbol{A})$$

Table 3. Diabetes study.

Outcome	Effect	Par.	ML	PL_P
LDL-chol.	Int. 1	$\xi_{1,01}$	−1.076 (0.108)	−1.073 (0.107)
Targets	Int. 2	$\xi_{1,02}$	0.155 (0.105)	1.157 (0.106)
	Int. 3	$\xi_{1,03}$	1.257 (0.110)	1.258 (0.115)
	Time	$\xi_{1,1}$	1.025 (0.076)	1.025 (0.071)
	Diab. dur. $T_0/10$	$\xi_{1,2}$	0.213 (0.088)	0.216 (0.090)
	Gender	$\xi_{1,3}$	0.497 (0.110)	0.497 (0.110)
	Insulin	$\xi_{1,4}$	0.853 (0.150)	0.829 (0.153)
	RI sd.	$\sqrt{d_{11}}$	1.852 (0.089)	1.849 (0.085)
HbA1c	Int. 1	$\xi_{2,01}$	0.901 (0.110)	0.902 (0.110)
Targets	Int. 2	$\xi_{2,02}$	3.195 (0.139)	3.194 (0.141)
	Time	$\xi_{2,1}$	1.068 (0.081)	1.068 (0.077)
	Diab.dur. $T_0/10$	$\xi_{2,2}$	−0.532 (0.088)	−0.531 (0.090)
	Gender	$\xi_{2,3}$	−0.183 (0.114)	−0.181 (0.115)
	Insulin	$\xi_{2,4}$	−1.103 (0.140)	−1.108 (0.145)
	RI sd.	$\sqrt{d_{22}}$	1.923 (0.095)	1.920 (0.089)
SBP	Int. 1	$\xi_{3,01}$	−0.151 (0.092)	−0.150 (0.093)
Targets	Int. 2	$\xi_{3,02}$	1.416 (0.099)	1.417 (0.100)
	Int. 3	$\xi_{3,03}$	3.740 (0.134)	3.741 (0.137)
	Time	$\xi_{3,1}$	0.508 (0.067)	0.509 (0.066)
	Diab. dur. $T_0/10$	$\xi_{3,2}$	−0.098 (0.076)	−0.096 (0.076)
	Gender	$\xi_{3,3}$	0.185 (0.096)	0.185 (0.097)
	Insulin	$\xi_{3,4}$	0.193 (0.127)	0.178 (0.129)
	RI sd.	$\sqrt{d_{33}}$	1.620 (0.076)	1.619 (0.076)
LDL-chol. targets and HbA1c targets	Cov. RI's	d_{12}	0.472 (0.151)	0.459 (0.158)
LDL-chol. targets and SBP targets	Cov. RI's	d_{13}	0.525 (0.125)	0.513 (0.126)
HbA1c targets and SBP targets	Cov. RI's	d_{23}	0.505 (0.131)	0.493 (0.136)

Note: Estimates (standard errors) of the regression coefficients for estimation methods for the Q = 3 case. ML: maximization of full likelihood; PL_P: pairwise modeling.

with $\boldsymbol{J}$ a block diagonal matrix with diagonal blocks equal to

$$\boldsymbol{J}_{m,pp} = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial^2 l_{m,pi}}{\partial \boldsymbol{\theta}_{m,p} \partial \boldsymbol{\theta}'_{m,p}}\right)$$

where $m = 1, \ldots, M,\ p = 1, \ldots, L(L-1)/2$. $\boldsymbol{K}$ is also a block-diagonal matrix but with larger blocks $\boldsymbol{K}_{m,\bullet}$. Each $\boldsymbol{K}_{m,\bullet}$ block corresponds to the pairwise modeling within a partition. It implies that $\boldsymbol{K}_{m,\bullet}$ is a symmetric matrix containing blocks

$$\boldsymbol{K}_{m,pq} = \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial l_{m,pi}}{\partial \boldsymbol{\theta}_{m,p}} \frac{\partial l_{m,qi}}{\partial \boldsymbol{\theta}'_{m,q}}\right)$$

with $p, q = 1, \ldots, L(L-1)/2$.

Table 4. Diabetes study.

Outcome	Effect	Par.	PL_s	PL_{ps}
LDL-chol.	Int. 1	$\xi_{1,01}$	−1.063 (0.109)	−1.061 (0.110)
Targets	Int. 2	$\xi_{1,02}$	0.183 (0.107)	0.185 (0.109)
	Int. 3	$\xi_{1,03}$	1.291 (0.112)	1.292 (0.118)
	Time	$\xi_{1,1}$	1.025 (0.077)	1.025 (0.072)
	Diab. dur. $T_0/10$	$\xi_{1,2}$	0.198 (0.090)	0.201 (0.091)
	Gender	$\xi_{1,3}$	0.497 (0.111)	0.497 (0.112)
	Insulin	$\xi_{1,4}$	0.877 (0.153)	0.852 (0.156)
	RI sd.	$\sqrt{d_{11}}$	1.853 (0.090)	1.849 (0.087)
HbA1c	Int. 1	$\xi_{2,01}$	0.883 (0.111)	0.883 (0.111)
Targets	Int. 2	$\xi_{2,02}$	3.189 (0.140)	3.187 (0.141)
	Time	$\xi_{2,1}$	1.083 (0.082)	1.084 (0.078)
	Diab. dur. $T_0/10$	$\xi_{2,2}$	−0.513 (0.090)	−0.511 (0.091)
	Gender	$\xi_{2,3}$	−0.152 (0.114)	−0.150 (0.116)
	Insulin	$\xi_{2,4}$	−1.151 (0.141)	−1.157 (0.144)
	RI sd.	$\sqrt{d_{22}}$	1.912 (0.096)	1.910 (0.091)
SBP	Int. 1	$\xi_{3,01}$	−0.139 (0.093)	−0.139 (0.094)
Targets	Int. 2	$\xi_{3,02}$	1.432 (0.100)	1.432 (0.101)
	Int. 3	$\xi_{3,03}$	3.776 (0.136)	3.777 (1.139)
	Time	$\xi_{3,1}$	0.508 (0.067)	0.509 (0.066)
	Diab. dur. $T_0/10$	$\xi_{3,2}$	−0.115 (0.078)	−0.113 (0.078)
	Gender	$\xi_{3,3}$	0.184 (0.097)	0.185 (0.097)
	Insulin	$\xi_{3,4}$	0.206 (0.128)	0.192 (0.130)
	RI sd.	$\sqrt{d_{33}}$	1.620 (0.077)	1.619 (0.077)
LDL-chol. targets and HbA1c targets	Cov. RI's	d_{12}	0.491 (0.155)	0.478 (0.164)
LDL-chol. targets and SBP targets	Cov. RI's	d_{13}	0.527 (0.127)	0.516 (0.130)
HbA1c targets and SBP targets	Cov. RI's	d_{23}	0.511 (0.133)	0.498 (0.138)

Note: Estimates (standard errors) of the regression coefficients for estimation methods for the Q = 3 case. PL_s: partitioned samples (M = 5); PL_{ps}: combined modeling (M = 5).

For pairwise modeling of independent subsamples, the pieces consisted of pairs constructed within every subsample. They are analyzed separately and the fitting results combined into a single point and precision estimator using the appropriate pseudo-likelihood rules. We illustrate this with an example. Assume the data were divided into $M = 2$ independent subsamples, and then for each partition, we obtain $\hat{\boldsymbol{\theta}}_m^*$ and its covariance by using equation (3) with N replaced by n. Denote the results of the first step by $\hat{\boldsymbol{\theta}}_1^*$, $\hat{\boldsymbol{\theta}}_2^*$ with the corresponding covariance matrices $\sum(\hat{\boldsymbol{\theta}}_1^*)$, $\sum(\hat{\boldsymbol{\theta}}_2^*)$. Then, by applying equation (4), the overall estimator of the parameters equals

$$\hat{\boldsymbol{\theta}}^* = \frac{1}{2}(\hat{\boldsymbol{\theta}}_1^* + \hat{\boldsymbol{\theta}}_2^*).$$

Table 5. Diabetes study.

	ML			PL_p			PL_s			PL_{ps}		
LDL-chol. Targets	1.000			1.000			1.000			1.000		
HbA1C Targets	0.133	1.000		0.129	1.000		0.139	1.000		0.136	1.000	
SBP Targets	0.175	0.162	1.000	0.171	0.159	1.00	0.176	0.165	1.000	0.172	0.161	1.00

Note: Estimated correlation matrix for the random intercepts for the Q = 3 case. ML: maximization of full likelihood; PL_p: pairwise modeling; PL_s: partitioned samples (M = 5); PL_{ps}: combined modeling (M = 5).

Table 6. Diabetes study.

Outcome 1	Outcome 2	Outcome 3	Partition	Pair	CPU Time
ML: maximization of full likelihood					
LDL-Chol.targets	HbA1c Targets	SBP targets			0:07:13
PL_p : pairwise modeling, three parallel processes					
LDL-Chol. targets	HbA1C targets			1	0:01:12
LDL-Chol. targets	SBP targets			2	0:01:23
HbA1c targets	SBP targets			3	0:01:12
PL_s : partitioned samples (M = 5), five parallel processes					
LDL-Chol. targets	HbA1c targets	SBP targets	1		0:01:12
LDL-Chol. targets	HbA1c targets	SBP targets	2		0:01:21
LDL-Chol. targets	HbA1c targets	SBP targets	3		0:01:14
LDL-Chol. targets	HbA1c targets	SBP targets	4		0:01:18
LDL-Chol. targets	HbA1c targets	SBP targets	5		0:01:17
PL_{ps}: combined modeling (M = 5), 15 parallel processes					
LDL-Chol. targets	HbA1C targets		1	1	0:00:14
LDL-Chol. targets	SBP targets		1	2	0:00:17
HbA1c targets	SBP targets		1	3	0:00:15
LDL-Chol. targets	HbA1C targets		2	1	0:00:15
LDL-Chol. targets	SBP targets		2	2	0:00:20
HbA1c targets	SBP targets		2	3	0:00:17
LDL-Chol. targets	HbA1C targets		3	1	0:00:14
LDL-Chol. targets	SBP targets		3	2	0:00:17
HbA1c targets	SBP targets		3	3	0:00:16
LDL-Chol. targets	HbA1C targets		4	1	0:00:15
LDL-Chol. targets	SBP targets		4	2	0:00:18
HbA1c targets	SBP targets		4	3	0:00:16
LDL-Chol. targets	HbA1C targets		5	1	0:00:16
LDL-Chol. targets	SBP targets		5	2	0:00:19
HbA1c targets	SBP targets		5	3	0:00:17

Note: Computation time (hours:minutes:seconds) for different estimation methods for the Q = 3 case. ML: maximization of full likelihood; PL_p: pairwise modeling; PL_s: partitioned samples (M = 5); PL_{ps}: combined modeling (M = 5).

Table 7. Diabetes study.

Outcome	Effect	Par.	ML vs. PL_p	ML vs. PL_s	ML vs. PL_{ps}
LDL-chol.	Int. 1	ξ_{101}	1.008	0.974	0.969
Targets	Int. 2	ξ_{102}	0.967	0.967	0.922
	Int. 3	ξ_{103}	0.921	0.961	0.876
	Time	$\xi_{1,1}$	1.143	0.991	1.131
	Diab. dur. $T_0/10$	ξ_{12}	0.968	0.954	0.932
	Gender	ξ_{13}	0.987	0.976	0.958
	Insulin	$\xi_{1,4}$	0.967	0.968	0.929
	RI sd.	$\sqrt{d_{11}}$	1.091	0.966	1.041
HbA1c	Int. 1	ξ_{201}	1.001	0.990	0.991
Targets	Int. 2	ξ_{202}	0.974	0.988	0.963
	Time	$\xi_{2,1}$	1.122	0.978	1.088
	Diab. dur. $T_0/10$	ξ_{22}	0.958	0.950	0.943
	Gender	ξ_{23}	0.976	0.988	0.968
	Insulin	$\xi_{2,4}$	0.936	0.980	0.950
	RI sd.	$\sqrt{d_{22}}$	1.128	0.984	1.100
SBP	Int. 1	ξ_{301}	0.984	0.982	0.960
Targets	Int. 2	ξ_{302}	0.974	0.980	0.947
	Int. 3	ξ_{303}	0.953	0.971	0.923
	Time	$\xi_{3,1}$	1.032	0.995	1.027
	Diab. dur. $T_0/10$	ξ_{32}	1.014	0.953	0.946
	Gender	ξ_{33}	0.993	0.984	0.977
	Insulin	$\xi_{3,4}$	0.963	0.981	0.945
	RI sd.	$\sqrt{d_{33}}$	0.991	0.975	0.965
LDL-chol. targets and HbA1c targets	Cov. RI's	d_{12}	0.912	0.951	0.852
LDL-chol. targets and SBP targets	Cov. RI's	d_{13}	0.985	0.962	0.928
HbA1c targets and SBP targets	Cov. RI's	d_{23}	0.926	0.977	0.910

Note: ARE for maximum likelihood and pseudo-likelihood methods for the Q = 3 case. ML vs. PL_p= variance ML vs. variance PL_p; ML vs. PL_s= variance ML vs. variance PL_s (M = 5); ML vs. PL_{ps} = variance ML vs. variance PL_{ps} (M = 5).

Finally, the assymptotic covariance of $\hat{\boldsymbol{\theta}}^*$ can be constructed similar to equation (7)

$$\sum(\hat{\boldsymbol{\theta}}^*) = \frac{1}{N}\left[\frac{1}{2}\left\{\sum(\hat{\boldsymbol{\theta}}_1^*) + \sum(\hat{\boldsymbol{\theta}}_2^*)\right\}\right].$$

6 Analysis of diabetes study

As an illustration, we analyze the diabetes data introduced in Section 2. In this analysis, the experts defined *multiple targets* for HbA1C, LDL-cholestrol and SBD which are regarded as a joint trial endpoint. Denote the components of this joint endpoint as Y_{1ij}, Y_{2ij}, and Y_{3ij} with the subscripts i

Table 8. Diabetes study.

Outcome	Effect	Par.	ML	PL_P
LDL-chol.	Int. 1	ξ_{101}	−1.092 (0.112)	−1.089 (0.111)
Targets	Int. 2	ξ_{102}	0.167 (0.109)	0.169 (0.110)
	Int. 3	ξ_{103}	1.293 (0.114)	1.295 (0.119)
	Time	$\xi_{1,1}$	1.044 (0.077)	1.044 (0.073)
	Diab.dur. $T_0/10$	ξ_{12}	0.220 (0.091)	0.224 (0.093)
	Gender	ξ_{13}	0.513 (0.114)	0.513 (0.114)
	Insulin	$\xi_{1,4}$	0.879 (0.155)	0.854 (0.158)
	RI sd.	$\sqrt{d_{11}}$	1.956 (0.097)	1.954 (0.099)
HbA1c	Int. 1	ξ_{201}	0.964 (0.117)	0.965 (0.118)
Targets	Int. 2	ξ_{202}	3.331 (0.149)	3.331 (0.155)
	Time	$\xi_{2,1}$	1.095 (0.083)	1.095 (0.079)
	Diab.dur. $T_0/10$	ξ_{22}	−0.565 (0.093)	−0.564 (0.095)
	Gender	ξ_{23}	−0.190 (0.120)	−0.188 (0.120)
	Insulin	$\xi_{2,4}$	−1.114 (0.146)	−1.120 (0.151)
	RI sd.	$\sqrt{d_{22}}$	2.077 (0.105)	2.076 (0.107)
SBP	Int. 1	ξ_{301}	−0.143 (0.094)	−0.142 (0.095)
Targets	Int. 2	ξ_{302}	1.445 (0.101)	1.446 (0.103)
	Int. 3	ξ_{303}	3.800 (0.137)	3.797 (0.143)
	Time	$\xi_{3,1}$	0.514 (0.067)	0.515 (0.067)
	Diab. dur. $T_0/10$	ξ_{32}	−0.102 (0.078)	−0.099 (0.077)
	Gender	ξ_{33}	0.188 (0.099)	0.189 (0.098)
	Insulin	$\xi_{3,4}$	0.201 (0.129)	0.186 (0.132)
	RI sd.	$\sqrt{d_{33}}$	1.680 (0.080)	1.680 (0.084)
LDL-chol. targets and HbA1c targets	Cov. RI's	d_{12}	0.508 (0.164)	0.491 (0.168)
LDL-chol. targets and SBP targets	Cov. RI's	d_{13}	0.552 (0.133)	0.539 (0.133)
HbA1c targets and SBP targets	Cov. RI's	d_{23}	0.536 (0.142)	0.522 (0.145)

Note: Estimates (standard errors) of the regression coefficients for estimation methods for the Q = 15 case. ML: maximization of full likelihood; PL_P: pairwise modeling.

and j indicating a measurement for subject i ($i = 1, \ldots, 2259$) at occasion j ($j = 1, 2$). Since every component was defined on an ordinal scale, we can specify for every one of them a univariate POMM. For simplicity and without loss of generality, we will use the same covariate structure. More specifically, for the lth outcome ($l = 1, 2, 3$) with R categories on an ordinal scale, we assume the following model

$$\text{logit}\,[P(Y_{lij} \leq r)] = \xi_{l,0r} + \xi_{l,1} t_{ij} + \xi_{l,2} X_{1,i} + \xi_{l,3} X_{2,i} + \xi_{l,4} X_{3,ij},$$

where $r = 1, \ldots, (R-1)$, t_{ij} is the time point at which the outcome is measured, i.e. $t_{ij} = 0$ or 1. For the corresponding conditional models, in order to capture the association between the measurements within the same subject for a certain response, three random intercepts will be included: b_{1i}, b_{2i}, and b_{3i}.

Table 9. Diabetes study.

Outcome	Effect	Par.	PL_s	PL_{ps}
LDL-chol.	Int. 1	$\xi_{1,01}$	−1.077 (0.113)	−1.075 (0.113)
Targets	Int. 2	$\xi_{1,02}$	0.201 (0.111)	0.202 (0.113)
	Int. 3	$\xi_{1,03}$	1.335 (0.117)	1.336 (0.123)
	Time	$\xi_{1,1}$	1.045 (0.077)	1.044 (0.074)
	Diab. dur. $T_0/10$	$\xi_{1,2}$	0.206 (0.094)	0.209 (0.095)
	Gender	$\xi_{1,3}$	0.513 (0.116)	0.513 (0.116)
	Insulin	$\xi_{1,4}$	0.902 (0.158)	0.876 (0.161)
	RI sd.	$\sqrt{d_{11}}$	1.965 (0.098)	1.960 (0.102)
HbA1c	Int. 1	$\xi_{2,01}$	0.943 (0.117)	0.944 (0.118)
Targets	Int. 2	$\xi_{2,02}$	3.323 (0.149)	3.323 (0.156)
	Time	$\xi_{2,1}$	1.111 (0.084)	1.112 (0.081)
	Diab. dur. $T_0/10$	$\xi_{2,2}$	−0.544 (0.096)	−0.543 (0.096)
	Gender	$\xi_{2,3}$	−0.156 (0.121)	−0.155 (0.121)
	Insulin	$\xi_{2,4}$	−1.166 (0.148)	−1.172 (0.150)
	RI sd.	$\sqrt{d_{22}}$	2.066 (0.106)	2.066 (0.108)
SBP	Int. 1	$\xi_{3,01}$	−0.131 (0.095)	−0.131 (0.096)
Targets	Int. 2	$\xi_{3,02}$	1.462 (0.102)	1.463 (0.105)
	Int. 3	$\xi_{3,03}$	3.834 (0.140)	3.836 (0.146)
	Time	$\xi_{3,1}$	0.513 (0.068)	0.515 (0.067)
	Diab. dur. $T_0/10$	$\xi_{3,2}$	−0.120 (0.080)	−0.118 (0.080)
	Gender	$\xi_{3,3}$	0.187 (0.099)	0.189 (0.099)
	Insulin	$\xi_{3,4}$	0.217 (0.131)	0.203 (0.133)
	RI sd.	$\sqrt{d_{33}}$	1.682 (0.081)	1.682 (0.085)
LDL-chol. targets and HbA1c targets	Cov. RI's	d_{12}	0.524 (0.169)	0.508 (0.175)
LDL-chol. targets and SBP targets	Cov. RI's	d_{13}	0.553 (0.136)	0.541 (0.137)
HbA1c targets and SBP targets	Cov. RI's	d_{23}	0.544 (0.144)	0.530 (0.147)

Note: Estimates (standard errors) of the regression coefficients for estimation methods for the Q = 15 case. PL_s: partitioned samples (M = 5); PL_{ps}: combined modeling (M = 5).

Table 10. Diabetes study.

	ML			PL_p			PL_s			PL_{ps}		
LDL-chol. targets	1.000			1.000			1.000			1.000		
HbA1C targets	0.125	1.000		0.121	1.000		0.129	1.000		0.126	1.000	
SBP targets	0.168	0.154	1.000	0.164	0.150	1.00	0.167	0.156	1.000	0.164	0.153	1.00

Note: Estimated correlation matrix for the random intercepts for the Q = 15 case. ML: maximization of full likelihood; PL_p: pairwise modeling; PL_s: partitioned samples (M = 5); PL_{ps}: combined modeling (M = 5).

Table 11. Diabetes study.

Outcome 1	Outcome 2	Outcome 3	Partition	Pair	CPU Time
ML: maximization of full likelihood					
LDL-Chol.Targets	HbA1c Targets	SBP targets			10:02:42
PL_p: pairwise modeling, three parallel processes					
LDL-Chol. targets	HbA1C Targets			1	0:16:01
LDL-Chol. targets	SBP Targets			2	0:20:04
HbA1c targets	SBP Targets			3	0:16:52
PL_s: partitioned samples, M = 5, five parallel processes					
LDL-Chol. targets	HbA1c targets	SBP targets	1		2:08:40
LDL-Chol. targets	HbA1c targets	SBP targets	2		2:31:20
LDL-Chol. targets	HbA1c targets	SBP targets	3		2:14:29
LDL-Chol. targets	HbA1c targets	SBP targets	4		2:23:29
LDL-Chol. targets	HbA1c targets	SBP targets	5		2:24:58
PL_{ps}: combined modeling, M = 5, 15 parallel processes					
LDL-Chol. targets	HbA1C targets		1	1	0:03:09
LDL-Chol. targets	SBP targets		1	2	0:04:12
HbA1c targets	SBP targets		1	3	0:03:25
LDL-Chol. targets	HbA1C targets		2	1	0:03:30
LDL-Chol. targets	SBP targets		2	2	0:04:11
HbA1c targets	SBP targets		2	3	0:03:40
LDL-Chol. targets	HbA1C targets		3	1	0:02:58
LDL-Chol. targets	SBP targets		3	2	0:04:11
HbA1c targets	SBP targets		3	3	0:03:34
LDL-Chol. targets	HbA1C targets		4	1	0:03:23
LDL-Chol. targets	SBP targets		4	2	0:04:07
HbA1c targets	SBP targets		4	3	0:03:58
LDL-Chol. targets	HbA1C targets		5	1	0:03:30
LDL-Chol. targets	SBP targets		5	2	0:04:17
HbA1c targets	SBP targets		5	3	0:03:25

Note: Computation time (hours:minutes:seconds) for different estimation methods for the Q = 15 case. ML: maximization of full likelihood; PL_p: pairwise modeling; PL_s: partitioned samples (M = 5); PL_{ps}: combined modeling (M = 5).

In order to capture the correlation between three responses, an assumption about the distribution of the vector of random effects $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i})'$ should be made. For example, to apply the model, the following assumption can be specified

$$\boldsymbol{b}_i \sim N(\mathbf{0}, D) \tag{8}$$

where D is a covariance matrix of random effects with elements d_{uv} (u,v = 1,2,3).

Four analyses are performed: (1) the classical fitting of joint models by maximizing the full joint likelihood (ML) and then the three alternative pseudo-likelihood methods, (2) pairwise modeling (PL_p), (3) modeling using partitioned independent subsamples (PL_s), and (4) pairwise modeling of independent subsamples (PL_{ps}).

Parameter estimates of this model can be obtained by maximizing the full likelihood, based on subject specific contributions from equation (2), after integrating out the random effects by using

Table 12. Diabetes study.

Outcome	Effect	Par.	ML vs. PL_p	ML vs. PL_s	ML vs. PL_{ps}
LDL-chol.	Int. 1	$\xi_{1\beta 1}$	1.013	0.967	0.967
Targets	Int. 2	$\xi_{1\beta 2}$	0.972	0.959	0.917
	Int. 3	$\xi_{1\beta 3}$	0.916	0.953	0.861
	Time	$\xi_{1,1}$	1.113	0.991	1.099
	Diab. dur. $T_0/10$	ξ_{12}	0.978	0.949	0.937
	Gender	ξ_{13}	0.996	0.969	0.961
	Insulin	ξ_{14}	0.969	0.961	0.924
	RI sd.	$\sqrt{d_{11}}$	0.948	0.962	0.900
HbA1c	Int. 1	$\xi_{2\beta 1}$	0.986	0.991	0.977
Targets	Int. 2	$\xi_{2\beta 2}$	0.920	0.988	0.909
	Time	$\xi_{2,1}$	1.082	0.976	1.045
	Diab. dur. $T_0/10$	ξ_{22}	0.967	0.949	0.949
	Gender	ξ_{23}	0.994	0.987	0.983
	Insulin	ξ_{24}	0.936	0.979	0.950
	RI sd.	$\sqrt{d_{22}}$	0.975	0.985	0.950
SBP	Int. 1	$\xi_{3\beta 1}$	0.988	0.980	0.962
Targets	Int. 2	$\xi_{3\beta 2}$	0.963	0.978	0.935
	Int. 3	$\xi_{3\beta 3}$	0.923	0.969	0.891
	Time	$\xi_{3,1}$	1.015	0.995	1.010
	Diab. dur. $T_0/10$	ξ_{32}	1.020	0.951	0.948
	Gender	ξ_{33}	1.002	0.981	0.982
	Insulin	ξ_{34}	0.967	0.979	0.945
	RI sd.	$\sqrt{d_{33}}$	0.909	0.972	0.880
LDL-chol. Targets and HbA1c targets	Cov. RI's	d_{12}	0.957	0.940	0.882
LDL-chol. targets and SBP targets	Cov. RI's	d_{13}	1.006	0.951	0.939
HbA1c targets and SBP targets	Cov. RI's	d_{23}	0.952	0.974	0.933

Note: ARE for maximum likelihood and pseudo-likelihood methods for the Q = 15 case. ML vs. PL_p = variance ML vs. variance PL_p; ML vs. PL_s = variance ML vs. Variance PL_s (M = 5); ML vs. PL_{ps} = variance ML vs. variance PL_{ps} (M = 5).

approximate methods (e.g. possibly adaptive, Gaussian quadrature). Because of the low number of responses (here, only three), this method is computationally intensive but still feasible. It was implemented with the SAS procedure NLMIXED, where for illustration's sake, we took $Q = 3$ quadrature points in the approximation. The results for the estimates and the standard errors are listed in Table 3 under ML. It should be noted that, as the full likelihood is maximized, the asymptotic covariance of the parameter estimators are derived from the inverse of the Fisher's information matrix $\boldsymbol{I}^{-1}$. It implies that the variance reaches the lower Cramér-Rao bound and, hence, the obtained estimator is a Cramér-Rao efficient estimator.

The three alternative methods are based on the maximization of the pseudo-likelihood function. The theory behind these methods is decribed in Section 5. First, we will apply pairwise fitting (PL_p in Table 3). For our case study, the information from three bivariate POMM models should be combined. For $PL_{s,}$ our second pseudo-likelihood method, we split the data into M independent

Table 13. Simulation study for two time-points with *N Subjects* = 1000, *N random samples* = 1000: MSE and interval coverage for different estimation methods, for the Q = 3 case.

			ML		PL_p	
Outcome	Effect	Par.	10^2·MSE	Int.Cov.	10^2·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{101} = -0.740$	1.502	0.955	1.502	0.958
Targets	Int. 2	$\xi_{102} = 0.480$	1.518	0.942	1.518	0.941
	Int. 3	$\xi_{103} = 1.580$	1.848	0.930	1.851	0.928
	Time	$\xi_{1,1} = 1.050$	0.966	0.961	0.966	0.961
	Gender	$\xi_{12} = 0.450$	2.404	0.950	2.404	0.951
	RI sd.	$\sqrt{d_{11}} = 1.880$	1.853	0.893	1.881	0.868
HbA1c	Int. 1	$\xi_{201} = 0.270$	1.897	0.936	1.903	0.938
Targets	Int. 2	$\xi_{202} = 2.610$	3.487	0.906	3.525	0.902
	Time	$\xi_{2,1} = 1.040$	1.332	0.946	1.332	0.939
	Gender	$\xi_{22} = -0.090$	3.001	0.945	2.997	0.946
	RI sd.	$\sqrt{d_{22}} = 2.150$	3.407	0.830	3.484	0.787
SBP	Int. 1	$\xi_{301} = -0.190$	1.175	0.963	1.171	0.963
Targets	Int. 2	$\xi_{302} = 1.380$	1.317	0.957	1.317	0.955
	Int. 3	$\xi_{303} = 3.700$	2.908	0.954	2.907	0.951
	Time	$\xi_{3,1} = 0.510$	0.927	0.949	0.926	0.948
	Gender	$\xi_{32} = 0.190$	1.934	0.959	1.923	0.962
	RI sd.	$\sqrt{d_{33}} = 1.620$	1.219	0.941	1.257	0.920
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	4.876	0.931	4.819	0.940
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	3.585	0.920	3.606	0.920
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	4.546	0.932	4.558	0.937

ML: maximization of full likelihood; PL_p: pairwise modeling.

subsamples, with $M = 5$ (see PL_s in Table 4). Finally, we have our new, combined method of pairwise modeling of independent subsamples (PL_{ps} in Table 4). As before, $M = 5$ partitions were used. For PL_s and PL_{ps}, we have chosen to divide the data using a completely randomized design at subject level. While this is not strictly necessary, it can be sensible in many applications.

Note that there are four patients in the data with only SBP values and no values for LDL-Cholesterol and HbA1C. For these patients, for PL_p and PL_{ps} methods, the subjects-specific information for the LDL-cholestrol and HbA1C target pair will be set equal to zero in $\boldsymbol{K}$ and $\boldsymbol{J}$. Also, to integrate out the random effects in each submodel of the three pseudo-likelihood methods, we will use $Q = 3$ as in the full model. All analyses have been performed with SAS procedures NLMIXED and IML (version 9.3).

Several observations can be made. First, as the subsamples in PL_s and PL_{ps} are not all of equal size, the data were divided into roughly equal portions. For these two methods, the weights in matrix $\boldsymbol{A}$ of all subsamples are made equal. This is not fully optimal, but as observed by Molenberghs et al.,[17] this approximation does not affect the validity of the patitioned method and can only slightly affect the efficiency.

In general, the estimation methods of the parameters, based on the earlier defined pseudo-likelihood, should be assessed as valid ones. Except for a few cases (e.g. for some intercepts of

Table 14. Simulation study for two time-points with *N Subjects* = 1000, *N random samples* = 1000: MSE and interval coverage for different estimation methods, for the Q = 3 case. PL_s, M = 5: partitioned samples, M = 5; PL_s, M = 10: partitioned samples, M = 10.

			PL_s, M = 5		PL_s, M = 10	
Outcome	Effect	Par.	10^2·MSE	Int.Cov.	10^2·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{1,01} = -0.740$	1.555	0.953	1.633	0.878
Targets	Int. 2	$\xi_{1,02} = 0.480$	1.545	0.942	1.585	0.874
	Int. 3	$\xi_{1,03} = 1.580$	1.837	0.931	1.881	0.865
	Time	$\xi_{1,1} = 1.050$	0.991	0.965	1.039	0.884
	Gender	$\xi_{1,2} = 0.450$	2.468	0.952	2.542	0.879
	RI sd.	$\sqrt{d_{11}} = 1.880$	1.801	0.899	1.777	0.839
HbA1c	Int. 1	$\xi_{2,01} = 0.270$	1.952	0.941	2.023	0.868
Targets	Int. 2	$\xi_{2,02} = 2.610$	3.275	0.925	3.244	0.872
	Time	$\xi_{2,1} = 1.040$	1.368	0.947	1.443	0.872
	Gender	$\xi_{2,2} = -0.090$	3.126	0.951	3.273	0.877
	RI sd.	$\sqrt{d_{22}} = 2.150$	3.169	0.845	2.940	0.806
SBP	Int. 1	$\xi_{3,01} = -0.190$	1.198	0.960	1.240	0.883
Targets	Int. 2	$\xi_{3,02} = 1.380$	1.341	0.957	1.431	0.879
	Int. 3	$\xi_{3,03} = 3.700$	3.180	0.959	4.229	0.869
	Time	$\xi_{3,1} = 0.510$	0.944	0.943	0.984	0.869
	Gender	$\xi_{3,2} = 0.190$	2.000	0.955	2.089	0.880
	RI sd.	$\sqrt{d_{33}} = 1.620$	1.229	0.948	1.264	0.868
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	5.107	0.936	5.429	0.873
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	3.607	0.933	3.671	0.865
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	4.670	0.939	4.865	0.871

HbA1C targets and also some non-repeated measured covariates), where estimates for ML method slightly differ from those of PL_s and PL_{ps} methods, all pseudo-likelihood methods yield approximations that are very similar to those of the ML method. The standard errors of the PL methods are also approximately equal to those of the ML method, with the largest difference for the estimates of the covariances between the random intercepts, i.e. d_{12}, d_{13}, d_{23}. Those of the PL_{ps} method deviate the most, and in addition also the standard error of $\xi_{1,03}$, a category-specific intercept. To quantify these differences, we use the Asymptotic Relative Efficiency (ARE): the ratio of the variance obtained with the maximum likelihood method and the one obtained with one of the pseudo-likelihood methods. We summarize these ARE's for all estimates in Appendix 3, Table 7. From this table, we conclude that the majority of the parameters have an ARE larger than 95%, but for some cases it shrinks to 85–87% for the PL_{ps} method. In summary, when fitting the model with one of the alternative pseudo-likelihood methods, we have almost no loss in efficiency for the main effects except for some intercepts or some covariance parameters where a very small loss is observed.

As an additional aim of our analysis, we estimate the strength of the association between three outcomes. Table 5 presents the correlations obtained from the fitted covariance matrix defined in equation (8). PL_s and PL_{ps} methods estimate the correlations closer to the those of the full likelihood

Table 15. Simulation study for two time-points with *N Subjects* = 1000, *N random samples* = 1000: MSE and interval coverage for different estimation methods, for the Q = 3 case. $PL_{ps,}$ M = 5: combined modeling, M = 5; $PL_{ps,}$ M = 10: combined modeling, M = 10.

			$PL_{ps,}$ M = 5		$PL_{ps,}$ M = 10	
Outcome	Effect	Par.	10^2·MSE	Int.Cov.	10^2·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{101} = -0.740$	1.554	0.955	1.639	0.953
Targets	Int. 2	$\xi_{102} = 0.480$	1.544	0.946	1.594	0.947
	Int. 3	$\xi_{103} = 1.580$	1.836	0.930	1.886	0.931
	Time	$\xi_{1,1} = 1.050$	0.991	0.960	1.038	0.958
	Gender	$\xi_{12} = 0.450$	2.467	0.953	2.552	0.954
	RI sd.	$\sqrt{d_{11}} = 1.880$	1.827	0.879	1.793	0.887
HbA1c	Int. 1	$\xi_{201} = 0.270$	1.957	0.943	2.025	0.948
Targets	Int. 2	$\xi_{202} = 2.610$	3.304	0.915	3.260	0.937
	Time	$\xi_{2,1} = 1.040$	1.368	0.936	1.435	0.935
	Gender	$\xi_{22} = -0.090$	3.117	0.952	3.261	0.953
	RI sd.	$\sqrt{d_{22}} = 2.150$	3.261	0.810	3.045	0.847
SBP	Int. 1	$\xi_{301} = -0.190$	1.193	0.959	1.235	0.959
Targets	Int. 2	$\xi_{302} = 1.380$	1.332	0.955	1.408	0.951
	Int. 3	$\xi_{303} = 3.700$	3.127	0.954	4.066	0.936
	Time	$\xi_{3,1} = 0.510$	0.943	0.944	0.982	0.939
	Gender	$\xi_{32} = 0.190$	1.983	0.959	2.068	0.957
	RI sd.	$\sqrt{d_{33}} = 1.620$	1.270	0.931	1.300	0.933
LDL-chol. Targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	5.054	0.945	5.355	0.949
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	3.623	0.928	3.678	0.940
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	4.660	0.940	4.820	0.944

method than PL_p method. This is a quite logical result because for PL_p the correlations were estimated only once: one for each pair.

In practice, once the model is fitted, it is used for hypothesis testing. Because of the close connection between pseudo-likelihood and likelihood, inferences such as the Wald test, pseudo-score test and pseudo-likelihood test have been developed (see Molenberghs and Verbeke[13] chap. 9 and Geys et al.[26]). Suppose one is interested in testing whether there is a significant joint evolution over time for all outcomes. In this case, the asymptotic Wald test can be applied. For the ML method, this test returns a value of 408.53 for the chi-square statistic. Using the alternative distribution based on pseudo-likelihood, the same test can be performed also for the PL methods: for $PL_{p,}$ PL_s and PL_{ps}, the chi-square statistics are equal to 409.53, 407.86, and 406.60, respectively. Obviously, the values of the Wald statistic for the PL methods are very close to that of the ML method and they all correspond to $p < 0.0001$. Hence, we reach the same conclusion for the ML and the three PL methods: there is a significant joint evolution for all outcomes.

The most important advantage of replacing the full likelihood method by an alternative pseudo-likelihood method is the gain in computation time. As all methods (ML, $PL_{p,}$ $PL_{s,}$ and $PL_{ps)}$ were applied using the same computer platform, a similar way of programming, and the same starting

Table 16. Simulation study for two time-points with *N Subjects* = 1000, *N random samples* = 1000.

Estimation method	CPU Time on HPC	Factor for CPU: PC vs. HPC	CPU Time on PC
ML	0:00:51	1.53	0:01:18
PL_p	0:00:11	1.32	0:00:14
PL_s, M = 5	0:00:11	1.54	0:00:17
PL_s, M = 10	0:00:07	1.55	0:00:11
PL_{ps}, M = 5	0:00:02	1.41	0:00:04
PL_{ps}, M = 10	0:00:01	1.41	0:00:02

Note: Average computation time (hours:minutes:seconds) for the longest subprocess for different estimation methods, the Q = 3 case. ML: maximization of full likelihood; PL_p: pairwise modeling; PL_s, M = 5: partitioned samples, M = 5; PL_s, M = 10: partitioned samples, M = 10; PL_{ps}, M = 5: combined modeling, M = 5; PL_{ps}, M = 10: combined modeling, M = 10.

Table 17. Simulation study for different number of time-points with *N Subjects* = 10,000, *N random samples* = 200: MSE and interval coverage for the ML estimation method, for the Q = 3 case.

			ML					
			Two time-points		10 time-points		20 time-points	
Outcome	Effect	Par.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{101} = -0.740$	1.499	0.955	0.951	0.980	0.819	0.955
Targets	Int. 2	$\xi_{102} = 0.480$	1.838	0.945	0.978	0.970	0.820	0.965
	Int. 3	$\xi_{103} = 1.580$	3.154	0.835	1.055	0.965	0.814	0.955
	Time	$\xi_{1,1} = 1.050$	1.212	0.940	0.445	0.980	0.298	0.955
	Gender	$\xi_{12} = 0.450$	2.211	0.970	1.561	0.950	1.447	0.945
	RI sd.	$\sqrt{d_{11}} = 1.880$	6.882	0.440	0.567	0.840	0.300	0.920
HbA1c	Int. 1	$\xi_{201} = 0.270$	2.686	0.875	1.563	0.925	1.217	0.945
Targets	Int. 2	$\xi_{202} = 2.610$	11.62	0.575	1.806	0.925	1.230	0.945
	Time	$\xi_{2,1} = 1.040$	1.832	0.885	0.867	0.955	0.386	0.930
	Gender	$\xi_{22} = -0.090$	2.345	0.985	2.097	0.960	1.963	0.965
	RI sd.	$\sqrt{d_{22}} = 2.150$	19.62	0.085	1.434	0.675	0.700	0.835
SBP	Int. 1	$\xi_{301} = -0.190$	1.507	0.910	0.829	0.950	0.646	0.955
Targets	Int. 2	$\xi_{302} = 1.380$	1.734	0.925	0.826	0.960	0.619	0.965
	Int. 3	$\xi_{303} = 3.700$	4.061	0.915	1.059	0.950	0.770	0.940
	Time	$\xi_{3,1} = 0.510$	0.934	0.955	0.526	0.965	0.245	0.950
	Gender	$\xi_{32} = 0.190$	1.820	0.955	1.187	0.965	1.076	0.955
	RI sd.	$\sqrt{d_{33}} = 1.620$	2.932	0.755	0.341	0.875	0.245	0.900
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	7.105	0.895	2.564	0.940	2.340	0.940
LDL-chol. targets and SBP Targets	Cov. RI's	$d_{13} = 1.218$	5.247	0.840	1.521	0.965	1.242	0.955
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	8.650	0.805	2.272	0.925	1.804	0.940

Table 18. Simulation study for different number of time-points with *N Subjects* = 10,000, *N random samples* = 200: MSE and interval coverage for the PL_p estimation method, for the $Q = 3$ case.

			PL_p					
			Two time-points		10 time-points		20 time-points	
Outcome	Effect	Par.	$10^3 \cdot$MSE	Int.Cov.	$10^3 \cdot$MSE	Int.Cov.	$10^3 \cdot$MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{101} = -0.740$	1.501	0.955	0.952	0.975	0.834	0.960
Targets	Int. 2	$\xi_{102} = 0.480$	1.862	0.940	0.979	0.970	0.834	0.965
	Int. 3	$\xi_{103} = 1.580$	3.223	0.840	1.058	0.965	0.831	0.955
	Time	$\xi_{1,1} = 1.050$	1.216	0.940	0.446	0.980	0.299	0.955
	Gender	$\xi_{12} = 0.450$	2.219	0.970	1.570	0.950	1.470	0.940
	RI sd.	$\sqrt{d_{11}} = 1.880$	7.232	0.380	0.592	0.825	0.308	0.920
HbA1c	Int. 1	$\xi_{201} = 0.270$	2.757	0.870	1.579	0.925	1.254	0.940
Targets	Int. 2	$\xi_{202} = 2.610$	12.00	0.540	1.829	0.925	1.271	0.945
	Time	$\xi_{2,1} = 1.040$	1.843	0.880	0.866	0.955	0.387	0.935
	Gender	$\xi_{22} = -0.090$	2.342	0.985	2.108	0.960	2.023	0.960
	RI sd.	$\sqrt{d_{22}} = 2.150$	20.47	0.055	1.524	0.640	0.730	0.815
SBP	Int. 1	$\xi_{301} = -0.190$	1.511	0.905	0.839	0.950	0.665	0.955
Targets	Int. 2	$\xi_{302} = 1.380$	1.820	0.910	0.836	0.960	0.640	0.955
	Int. 3	$\xi_{303} = 3.700$	4.350	0.870	1.072	0.945	0.794	0.935
	Time	$\xi_{3,1} = 0.510$	0.943	0.955	0.527	0.965	0.246	0.950
	Gender	$\xi_{32} = 0.190$	1.820	0.955	1.203	0.965	1.102	0.955
	RI sd.	$\sqrt{d_{33}} = 1.620$	3.379	0.685	0.360	0.865	0.250	0.895
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	6.839	0.900	2.544	0.945	2.359	0.930
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	5.426	0.825	1.519	0.965	1.246	0.955
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	9.179	0.795	2.286	0.935	1.808	0.940

values, we can fairly compare the computation times. The results are summarized in Table 6. The algorithm for the full likelihood method, ML, converged in 7 min 13 s. As with the pseudo-likelihood methods, the submodeling processes can be regarded as independent from one another and they can run in parallel on different computers. For example, for $PL_{p,}$ the computation time decreases to 1 min 23 s, i.e. the longest computation time among all parallel processes. A quite similar finding is obtained for PL_s : 1 min 21 s. But for PL_{ps}, we only need 20 s, a spectacular gain in speed. This clearly illustrates that the main advantage of the pairwise model fitting of partitioned data sets is that it allows fitting complex models to large data sets, which in some cases would not be feasible with standard model fitting procedures by maximum likelihood.

In addition, to investigate the validity of the classical ML method as well as that of the alternative PL methods, a series of simulations were performed. The selected scenario was similar to the diabetes study where in the univariate POMM model, for every response, only time and covariate gender were included. The simulations were performed for 1000 subjects measured longitudinally on two time-points with three outcomes and 1000 random samples; these were generated from the joint population of three ordinal variables with the covariance structure as discussed above. The correlations were set to the following magnitudes: 0.2, 0.4, and 0.4. For numerical integration, $Q = 3$ was used. For PL_s and

Table 19. Simulation study for different number of time-points with *N Subjects* = 10,000, *N random samples* = 200: MSE and interval coverage for the PL_s M = 5 estimation method, for the Q = 3 case.

			PL_s, M = 5					
			Two time-points		10 time-points		20 time-points	
Outcome	Effect	Par.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{1,0,1} = -0.740$	1.500	0.950	0.957	0.970	0.844	0.960
Targets	Int. 2	$\xi_{1,0,2} = 0.480$	1.830	0.940	0.983	0.970	0.844	0.965
	Int. 3	$\xi_{1,0,3} = 1.580$	3.057	0.845	1.059	0.965	0.840	0.955
	Time	$\xi_{1,1} = 1.050$	1.202	0.935	0.446	0.980	0.298	0.950
	Gender	$\xi_{1,2} = 0.450$	2.209	0.975	1.562	0.950	1.482	0.945
	RI sd.	$\sqrt{d_{11}} = 1.880$	6.808	0.430	0.596	0.820	0.317	0.900
HbA1c	Int. 1	$\xi_{2,0,1} = 0.270$	2.662	0.875	1.571	0.925	1.253	0.940
Targets	Int. 2	$\xi_{2,0,2} = 2.610$	11.05	0.590	1.804	0.925	1.263	0.945
	Time	$\xi_{2,1} = 1.040$	1.780	0.880	0.869	0.955	0.387	0.930
	Gender	$\xi_{2,2} = -0.090$	2.343	0.985	2.129	0.965	2.032	0.965
	RI sd.	$\sqrt{d_{22}} = 2.150$	19.24	0.085	1.477	0.650	0.728	0.815
SBP	Int. 1	$\xi_{3,0,1} = -0.190$	1.513	0.910	0.833	0.950	0.654	0.955
Targets	Int. 2	$\xi_{3,0,2} = 1.380$	1.706	0.940	0.829	0.965	0.627	0.970
	Int. 3	$\xi_{3,0,3} = 3.700$	3.839	0.915	1.056	0.950	0.776	0.940
	Time	$\xi_{3,1} = 0.510$	0.937	0.955	0.526	0.965	0.244	0.950
	Gender	$\xi_{3,2} = 0.190$	1.828	0.955	1.191	0.965	1.089	0.950
	RI sd.	$\sqrt{d_{33}} = 1.620$	2.908	0.760	0.356	0.870	0.255	0.890
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	7.022	0.895	2.560	0.940	2.340	0.940
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	5.158	0.840	1.538	0.965	1.253	0.960
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	8.372	0.810	2.273	0.930	1.809	0.945

PL_{ps}, the data were split into $M = 5$ and $M = 10$ independent subsamples. Further, the precision of each parameter for the ML and PL methods was assessed using mean square error (MSE) and interval coverage. Finally, all results are summarized in Tables 13–15 (see Appendix 5). When comparing outcomes of all methods, we observe that MSE is slightly higher for some category-specific intercepts, for the non-repeated measured covariate, and for some components of the covariance matrix. Almost for all methods, the interval coverage is high for all fixed effects, and also for the covariances of the random effects, but slightly lower for the standard deviations of the random effects. The exception is with the PL_s method with $M = 10$: here, the magnitude of the interval coverage is overall lower compared to other methods. The low coverage was due to the small sample size of only 100 subjects when the asymptotic distribution of parameter estimates could not be reached. Also, from the 1000 generated random samples, for PL_s given $M = 10$, 77 fitted models and for PL_{ps} given $M = 10$, 28 fitted models did not converge and an additional number of random samples was simulated to reach the required number of 1000.

Because the simulations were performed on HPC, in order to estimate the computation time required by the originally used computational platform, we re-ran several of them on that platform. The results are shown in Table 16 of Appendix 5. Again, the PL_{ps} method turns out to be the fastest.

Table 20. Simulation study for different number of time-points with *N Subjects* = 10,000, *N random samples* = 200: MSE and interval coverage for PL_s M = 10 estimation method, for the Q = 3 case.

			PL_s, M = 10					
			Two time-points		10 time-points		20 time-points	
Outcome	Effect	Par.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{1,0,1} = -0.740$	1.509	0.950	0.955	0.975	0.845	0.950
Targets	Int. 2	$\xi_{1,0,2} = 0.480$	1.823	0.940	0.980	0.970	0.845	0.965
	Int. 3	$\xi_{1,0,3} = 1.580$	2.956	0.845	1.054	0.965	0.842	0.955
	Time	$\xi_{1,1} = 1.050$	1.190	0.935	0.447	0.980	0.299	0.945
	Gender	$\xi_{1,2} = 0.450$	2.197	0.975	1.564	0.955	1.494	0.945
	RI sd.	$\sqrt{d_{11}} = 1.880$	6.672	0.465	0.628	0.805	0.336	0.895
HbA1c	Int. 1	$\xi_{2,0,1} = 0.270$	2.642	0.875	1.577	0.925	1.263	0.950
Targets	Int. 2	$\xi_{2,0,2} = 2.610$	10.38	0.620	1.797	0.925	1.271	0.945
	Time	$\xi_{2,1} = 1.040$	1.723	0.885	0.874	0.955	0.387	0.935
	Gender	$\xi_{2,2} = -0.090$	2.331	0.985	2.139	0.965	2.044	0.960
	RI sd.	$\sqrt{d_{22}} = 2.150$	18.79	0.085	1.539	0.635	0.770	0.800
SBP	Int. 1	$\xi_{3,0,1} = -0.190$	1.511	0.915	0.836	0.955	0.658	0.960
Targets	Int. 2	$\xi_{3,0,2} = 1.380$	1.666	0.940	0.831	0.960	0.631	0.970
	Int. 3	$\xi_{3,0,3} = 3.700$	3.613	0.925	1.061	0.950	0.783	0.940
	Time	$\xi_{3,1} = 0.510$	0.936	0.955	0.526	0.970	0.245	0.950
	Gender	$\xi_{3,2} = 0.190$	1.827	0.960	1.192	0.965	1.090	0.955
	RI sd.	$\sqrt{d_{33}} = 1.620$	2.891	0.765	0.374	0.855	0.267	0.880
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	6.917	0.895	2.594	0.940	2.350	0.940
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	4.992	0.845	1.550	0.965	1.252	0.960
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	8.090	0.830	2.305	0.935	1.834	0.935

To investigate in depth the said problem with interval coverage, a limited set of simulations in the same settings as before but now with a much larger number of subjects (10,000) was performed using HPC. For this case, only 200 random samples were generated. The results for all methods are listed in Tables 17–22 (see Appendix 5). For PL_s, $M = 10$ (Table 20, second column) we observe that the interval coverage is in general higher for the fixed effects when the subsamples include a larger number of subjects. However, for all methods, the interval coverage of the standard deviation of some random effects is dramatically low. When increasing the number of time-points over which the subject was measured, the interval coverage increased to a reasonably high value. Hence, in order to reach the asymptotic distribution of the estimates of all components of the covariance matrix, a larger number of time-points for the subject measurements are required.

7 Concluding remarks

Pairwise and partition pseudo-likelihood fitting were applied to multivariate joint proportional odds mixed models. In addition, a new combined method of pairwise modeling of subsamples was introduced. Partitioning here was done into independent subsamples, where each subsample

Table 21. Simulation study for different number of time-points with *N Subjects* = 10,000, *N random samples* = 200: MSE and interval coverage for PL_{ps} M = 5 estimation method, for the Q = 3 case.

			PL_{ps}, M = 5					
			Two time-points		10 time-points		20 time-points	
Outcome	Effect	Par.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{1,0,1} = -0.740$	1.501	0.955	0.955	0.970	0.843	0.960
Targets	Int. 2	$\xi_{1,0,2} = 0.480$	1.849	0.940	0.981	0.970	0.843	0.965
	Int. 3	$\xi_{1,0,3} = 1.580$	3.116	0.835	1.058	0.965	0.839	0.955
	Time	$\xi_{1,1} = 1.050$	1.205	0.935	0.446	0.980	0.299	0.950
	Gender	$\xi_{1,2} = 0.450$	2.216	0.975	1.566	0.950	1.481	0.945
	RI sd.	$\sqrt{d_{11}} = 1.880$	7.150	0.380	0.619	0.810	0.325	0.895
HbA1c	Int. 1	$\xi_{2,0,1} = 0.270$	2.736	0.870	1.579	0.925	1.262	0.940
Targets	Int. 2	$\xi_{2,0,2} = 2.610$	11.43	0.570	1.817	0.925	1.275	0.945
	Time	$\xi_{2,1} = 1.040$	1.788	0.880	0.871	0.955	0.387	0.940
	Gender	$\xi_{2,2} = -0.090$	2.340	0.985	2.124	0.960	2.042	0.965
	RI sd.	$\sqrt{d_{22}} = 2.150$	20.11	0.060	1.565	0.630	0.759	0.795
SBP	Int. 1	$\xi_{3,0,1} = -0.190$	1.516	0.915	0.836	0.955	0.664	0.955
Targets	Int. 2	$\xi_{3,0,2} = 1.380$	1.786	0.920	0.834	0.960	0.637	0.965
	Int. 3	$\xi_{3,0,3} = 3.700$	4.086	0.890	1.064	0.950	0.788	0.940
	Time	$\xi_{3,1} = 0.510$	0.945	0.950	0.526	0.965	0.245	0.950
	Gender	$\xi_{3,2} = 0.190$	1.826	0.955	1.203	0.960	1.105	0.955
	RI sd.	$\sqrt{d_{33}} = 1.620$	3.358	0.690	0.376	0.855	0.259	0.885
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	6.755	0.900	2.538	0.945	2.350	0.940
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	5.329	0.830	1.529	0.960	1.254	0.960
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	8.898	0.790	2.284	0.935	1.813	0.945

consists of an independent set of patients. The attraction of the methods is that various forms of partitioning can be done at the same time: (a) rearranging a large set of simultaneously measured longitudinal sequences into pairs (b) into independent subsets of subjects, (c) splitting the potentially long longitudinal sequences into shorter subsequences. Clearly, both (a) and (c) lead to dependent subsamples, for which general pseudo-likelihood inference can be used. Further, as investigated by Molenberghs et al.,[17] the splitting of long longitudinal sequences in (c) into shorter ones strongly depends on what is practically and numerically feasible. And longer subsequences lead to smaller losses in efficiency.

After comparing the three pseudo-likelihood methods with the one based on full likelihood on the diabetes data, the following observations were made. First, even for low numbers of quadrature points ($Q = 3$), the alternative PL methods yield valid estimates with high efficiency. The efficiency of the alternative methods was slightly lower for some estimates of the covariance between the random intercepts, in particular for the combined method. This was also the case for some category-specific intercepts.

The big advantage of the alternative methods is their gain in computation time over the full likelihood method. Even if it would be feasible to apply the full model on all response components,

Table 22. Simulation study for different number of time-points with *N Subjects* = 10,000, *N random samples* = 200: MSE and interval coverage for PL_{ps} M = 10 estimation method, for the Q = 3 case.

			$PL_{ps,}$ M=10					
			Two time-points		10 time-points		20 time-points	
Outcome	Effect	Par.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.	10^3·MSE	Int.Cov.
LDL-chol.	Int. 1	$\xi_{1,01} = -0.740$	1.509	0.950	0.954	0.970	0.843	0.955
Targets	Int. 2	$\xi_{1,02} = 0.480$	1.841	0.945	0.979	0.970	0.844	0.965
	Int. 3	$\xi_{1,03} = 1.580$	3.010	0.845	1.054	0.965	0.840	0.955
	Time	$\xi_{1,1} = 1.050$	1.192	0.935	0.448	0.980	0.299	0.950
	Gender	$\xi_{1,2} = 0.450$	2.202	0.975	1.567	0.955	1.488	0.940
	RI sd.	$\sqrt{d_{11}} = 1.880$	7.006	0.390	0.651	0.800	0.344	0.890
HbA1c	Int. 1	$\xi_{2,01} = 0.270$	2.713	0.870	1.584	0.925	1.270	0.945
Targets	Int. 2	$\xi_{2,02} = 2.610$	10.76	0.590	1.808	0.925	1.280	0.945
	Time	$\xi_{2,1} = 1.040$	1.732	0.885	0.875	0.955	0.387	0.935
	Gender	$\xi_{2,2} = -0.090$	2.325	0.985	2.134	0.965	2.048	0.955
	RI sd.	$\sqrt{d_{22}} = 2.150$	19.69	0.075	1.631	0.585	0.802	0.780
SBP	Int. 1	$\xi_{3,01} = -0.190$	1.514	0.915	0.838	0.955	0.669	0.955
Targets	Int. 2	$\xi_{3,02} = 1.380$	1.739	0.920	0.835	0.965	0.643	0.965
	Int. 3	$\xi_{3,03} = 3.700$	3.817	0.905	1.066	0.950	0.795	0.940
	Time	$\xi_{3,1} = 0.510$	0.943	0.950	0.526	0.965	0.245	0.945
	Gender	$\xi_{3,2} = 0.190$	1.824	0.960	1.201	0.960	1.108	0.950
	RI sd.	$\sqrt{d_{33}} = 1.620$	3.344	0.700	0.397	0.845	0.273	0.875
LDL-chol. targets and HbA1c targets	Cov. RI's	$d_{12} = 0.808$	6.647	0.895	2.563	0.950	2.361	0.940
LDL-chol. targets and SBP targets	Cov. RI's	$d_{13} = 1.218$	5.144	0.840	1.539	0.960	1.256	0.960
HbA1c targets and SBP targets	Cov. RI's	$d_{23} = 1.393$	8.605	0.810	2.321	0.925	1.839	0.940

one could still prefer to proceed with one of the pseudo-likelihood methods. Indeed, as our example shows for only three components, we can achieve a significant reduction in computation time: from 7 min 13 s to only 20 s using the combined approach. This was due to the fact that the submodels could run in parallel. When confronted with time restrictions, it is therefore recommended to consider pairwise fitting of independent partitions.

Currently, as large data sets are often collected and stored, the number of response outcomes increases rapidly. When fitting the full model becomes unrealistic, pseudo-likelihood methods could offer a solution. Also, as the subprocesses could run in a parallel, the required computation time becomes feasible. When modeling, the main interest is with the estimation and inference for the fixed effects, properly accounting for dependencies in the data. At the same time, the association parameters may be of, perhaps secondary, interest as well. The pairwise setting is the minimal one that allows identification of all the parameters. Evidently, one could consider tripels or higher tupels as well, but this will arguably lead to minimal increase of efficiency, while drastically increase computational burden.

If we increase the number of quadrature points to achieve a better approximation, it is not unexpected that the computation time will increase as well. For example, for $Q = 15$, the

implementation of the full likelihood model for the diabetes example will take more than 10 h, whereas the combined method needs only 4 min (see Appendix 4, Table 11). Hence, the gain in computation time becomes even more relevant. When inspecting Tables 8 and 9 for $Q = 15$, we observe a small loss in quality for a few estimates, similar to $Q = 3$. From Table 10, we can conclude that the estimates of the correlations are lower for Q = 15 than for Q = 3, but for the PLs and PLps methods they are still closer to the full likelihood method than for the PLp method.

The SAS code developed for the combined method is available from the authors' website.

From the simulation study, we can conclude that all PL methods yield quite a high precision for the fixed effects, when taking into account that the data are not split into small subsamples for the PL_s and PL_{ps} methods. To reach asymptotic normality for the parameter estimates of all components of the covariance matrix, a larger number of time-points for the subject measurements are required.

The longitudinal and hierarchical settings are similar to those of the meta-analysis. Hence, the developed methodology can be easily and effectively applied in meta-analysis.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) and the Flemish Supercomputer Project are gratefully acknowledged. We are also grateful to Mr Kris Bogaerts of I-Biostat for his expert advice.

References

1. Laird N and Ware J. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–974.
2. Breslow NE and Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993; **88**: 9–25.
3. Wolfinger R and O'Connell M. Generalized linear mixed models: A pseudo-likelihood approach. *J Stat Comput Simul* 1993; **48**: 233–243.
4. Engel B and Keen A. A simple approach for the analysis of generalized linear mixed models. *Stat Neerlandica* 1994; **48**: 1–22.
5. Agresti A and Lang JB. A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika* 1993; **80**: 527–534.
6. Hedeker D and Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics* 1994; **50**: 933–944.
7. Hartzel J, Agresti A and Caffo B. Multinomial logit random effects models. *Stat Model* 2001; **1**: 81–102.
8. Morrell C, Brant L, Sheng S, et al. Screening for prostate cancer using multivariate mixed-effects models. *J Appl Stat* 2012; **39**: 1151–1175.
9. Gueorguieva R. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Stat Model* 2001; **1**: 177–193.
10. Iddi S and Molenberghs G. A joint marginalized multilevel model for continuous and binary longitudinal outcomes. *J Appl Stat* 2012; **39**: 2413–2430.
11. Rizopoulos D. *Joint models for logitudinal and time-to-event data with applications in R*. Boca Raton, FL: Chapman & Hall/CRC, 2012.
12. Tsiatis A and Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* 2004; **14**: 809–834.
13. Molenberghs G and Verbeke G. *Models for discrete longitudinal data*. New York, NY: Springer, 2005.
14. Fieuws S and Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* 2006; **62**: 424–431.
15. Fieuws S, Verbeke G, Boen F, et al. High-dimensional multivariate mixed models for binary questionnaire data. *Appl Stat* 2006; **55**: 1–12.
16. Fieuws S, Verbeke G, Maes B, et al. Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics* 2008; **9**: 419–431.
17. Molenberghs G, Verbeke G and Iddi S. Pseudo-likelihood methodology for partitioned large and complex samples. *Stat Probab Lett* 2011; **81**: 892–901.
18. Vasdekis VGS, Rizopoulos D and Moustaki I. A weighted pairwise likelihood estimator for a class of latent variable models. *Biostatistics* 2014; **15**: 677–689.
19. Borgermans L, Goderis G, Van Den Broeke C, et al. Interdisciplinary diabetes care teams operating on the interface between primary and specialty care are associated with improved outcomes of care: Findings from the Leuven Diabetes Project. *BMC Health Service Res* 2009; **9**: 179.

20. Molenberghs G, Verbeke G and Demétrio C. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Anal* 2007; **13**: 513–531.
21. Molenberghs G, Verbeke G, Demétrio CGB, et al. A family of generalized linear models for repeated measures with normal and conjugate random effects. *Stat Sci* 2010; **25**: 325–347.
22. Pinheiro JC and Bates DM. Approximations to the log-likelihood function in the non-linear mixed-effects model. *J Computat Graph Stat* 1995; **4**: 12–35.
23. Pinheiro JC and Bates DM. *Mixed effects models in S and S-Plus*. New York, NY: Springer, 2000.
24. Godambe VP (ed.) *Estimating functions*. Oxford: Clarendon Press, 1991.
25. Arnold BC and Strauss D. Pseudolikelihood estimation: Some examples. *Sankhya B* 1991; **53**: 233–243.
26. Geys H, Molenberghs G and Ryan L. Pseudo-likelihood modelling of multivariate outcomes in developmental toxicology. *J Am Stat Assoc* 1999; **94**: 734–745.

Appendix

1. An asymptotic normal distribution of parameters in case of pairwise modeling
   An asymptotic multivariate distribution for $\hat{\boldsymbol{\theta}}$ is

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, \boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1})$$

   where $\boldsymbol{J}$ is a block diagonal matrix with blocks $\boldsymbol{J}_{pp}$ and $\boldsymbol{K}$ a symmetric matrix containing blocks $\boldsymbol{K}_{pq}$

$$\boldsymbol{J}_{pp} = -\frac{1}{N}\sum_{i=1}^{N} E\left(\frac{\partial^2 l_{pi}}{\partial\boldsymbol{\theta}_p \partial\boldsymbol{\theta}'_p}\right),$$

$$\boldsymbol{K}_{pq} = \frac{1}{N}\sum_{i=1}^{N} E\left(\frac{\partial l_{pi}}{\partial\boldsymbol{\theta}_p}\frac{\partial l_{qi}}{\partial\boldsymbol{\theta}'_q}\right),$$

   with $p, q = 1, \ldots, L(L-1)/2$.
2. An asymptotic normal distribution of parameters in case of partitioned samples
   In case of independent subsamples, blocks $\boldsymbol{J}_m$ and $\boldsymbol{K}_m$ $(m = 1, \ldots, M)$ can be regarded as a simplification of the blocks of Appendix 1. They are equal up to the sign

$$\boldsymbol{J}_m = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial^2 l_{mi}}{\partial\boldsymbol{\theta}_m \partial\boldsymbol{\theta}'_m}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial l_{mi}}{\partial\boldsymbol{\theta}_m}\frac{\partial l_{mi}}{\partial\boldsymbol{\theta}'_m}\right) = -\boldsymbol{K}_m.$$

The covariance of $\sqrt{N}\hat{\boldsymbol{\theta}}$ corresponds to the matrix product $\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1} = -\boldsymbol{J}^{-1}$. The asymptotic distribution for the overall averaged vector of parameter estimators can be expressed as follows

$$\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) = \sqrt{N}(\boldsymbol{A}'\hat{\boldsymbol{\theta}} - \boldsymbol{A}'\boldsymbol{\theta}) \sim N(\mathbf{0}, -M \cdot \boldsymbol{A}'\boldsymbol{J}^{-1}\boldsymbol{A})$$

with

$$\boldsymbol{A} = \frac{1}{M}(\boldsymbol{I}, \boldsymbol{I}, \ldots, \boldsymbol{I}).$$

3. Asymptotic relative efficency (ARE) for maximum likelihood and pseudo-likelihood methods with quadrature points $Q = 3$ in approximate integration
4. Results for $Q = 15$ quadrature points in approximate integration
5. Simulation study